



Full wwPDB X-ray Structure Validation Report ⓘ

Mar 5, 2026 – 02:32 PM UTC

PDB ID : 1KJG / pdb_00001kjg
Title : SUBSTRATE SHAPE DETERMINES SPECIFICITY OF RECOGNITION
RECOGNITION FOR HIV-1 PROTEASE: ANALYSIS OF CRYSTAL
STRUCTURES OF SIX SUBSTRATE COMPLEXES
Authors : Schiffer, C.A.
Deposited on : 2001-12-04
Resolution : 2.00 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4-5-2 with Phenix2.0
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

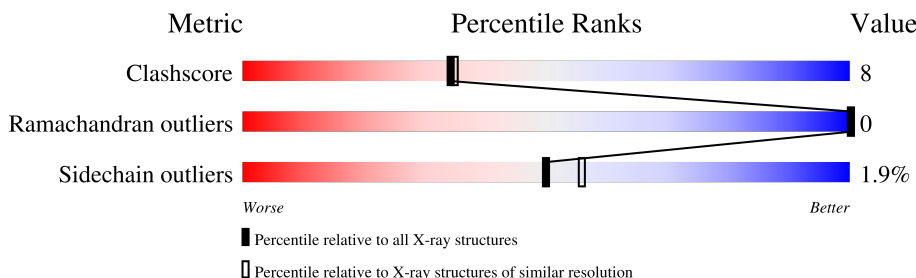
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.00 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	190562	11152 (2.00-2.00)
Ramachandran outliers	187476	11031 (2.00-2.00)
Sidechain outliers	187428	11029 (2.00-2.00)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	99	
1	B	99	
2	P	10	

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 1716 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called POL POLYPROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	99	Total	C	N	O	S	0	4	0
			772	495	132	141	4			
1	B	99	Total	C	N	O	S	0	4	0
			778	499	136	139	4			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	7	LYS	GLN	engineered mutation	UNP P03369
A	25	ASN	ASP	engineered mutation	UNP P03369
B	7	LYS	GLN	engineered mutation	UNP P03369
B	25	ASN	ASP	engineered mutation	UNP P03369

- Molecule 2 is a protein called GAG POLYPROTEIN.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	P	9	Total	C	N	O	0	0	0
			69	44	9	16			

- Molecule 3 is ACETATE ION (CCD ID: ACT) (formula: C₂H₃O₂).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			4	2	2		
3	A	1	Total	C	O	0	0
			4	2	2		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	37	Total	O	0	0
			37	37		
4	B	48	Total	O	0	0
			48	48		
4	P	4	Total	O	0	0
			4	4		

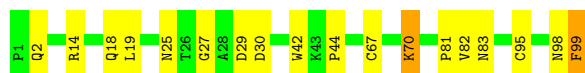
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

Note EDS was not executed.

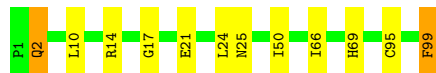
- Molecule 1: POL POLYPROTEIN

Chain A:  82% 16% .



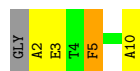
- Molecule 1: POL POLYPROTEIN

Chain B:  88% 10% .



- Molecule 2: GAG POLYPROTEIN

Chain P:  50% 30% 10% 10%



4 Data and refinement statistics

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	51.30Å 58.65Å 62.10Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	32.79 – 2.00	Depositor
% Data completeness (in resolution range)	98.6 (32.79-2.00)	Depositor
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	CNS	Depositor
R, R_{free}	0.184 , 0.226	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	1716	wwPDB-VP
Average B, all atoms (Å ²)	24.0	wwPDB-VP

5 Model quality [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: ACT

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\# Z > 5$	RMSZ	$\# Z > 5$
1	A	0.61	2/723 (0.3%)	0.86	0/980
1	B	0.57	2/725 (0.3%)	0.89	1/982 (0.1%)
2	P	0.93	0/70	1.02	1/93 (1.1%)
All	All	0.61	4/1518 (0.3%)	0.88	2/2055 (0.1%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	99	PHE	C-O	-8.94	1.05	1.23
1	B	99	PHE	C-O	-7.21	1.09	1.23
1	B	2	GLN	CD-OE1	5.18	1.33	1.23
1	A	2	GLN	CD-OE1	5.12	1.33	1.23

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	P	5	PHE	N-CA-C	5.61	118.48	110.68
1	B	25	ASN	N-CA-C	5.03	117.81	107.69

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	772	0	787	20	0
1	B	778	0	805	9	0
2	P	69	0	56	4	0
3	A	8	0	6	0	0
4	A	37	0	0	1	0
4	B	48	0	0	0	0
4	P	4	0	0	1	0
All	All	1716	0	1654	25	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 8.

All (25) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:81:PRO:HD2	1:B:50[1]:ILE:HD13	1.61	0.82
1:A:70:LYS:HB3	1:A:70:LYS:HZ3	1.55	0.72
1:A:98:ASN:HD21	1:B:2:GLN:HG3	1.55	0.72
1:A:25:ASN:ND2	1:A:27:GLY:H	1.97	0.62
1:A:98:ASN:ND2	1:B:2:GLN:HG3	2.16	0.59
1:A:81:PRO:HD2	1:B:50[1]:ILE:CD1	2.33	0.58
1:A:25:ASN:HD22	1:A:27:GLY:H	1.52	0.57
2:P:10:ALA:HB3	4:P:14:HOH:O	2.08	0.53
1:A:42:TRP:HZ3	1:A:44:PRO:HG3	1.75	0.52
1:B:10:LEU:HB3	1:B:21[2]:GLU:CD	2.36	0.50
2:P:3:GLU:HB3	2:P:5:PHE:CZ	2.46	0.50
1:A:30:ASP:OD2	2:P:2:ALA:HB3	2.12	0.49
1:A:42:TRP:CZ3	1:A:44:PRO:HG3	2.48	0.49
1:A:30:ASP:HB3	4:A:505:HOH:O	2.13	0.48
1:A:14:ARG:HA	1:A:18:GLN:O	2.14	0.48
1:A:25:ASN:HD22	1:A:25:ASN:C	2.21	0.48
1:A:99:PHE:HB3	1:B:95:CYS:HA	1.98	0.45
1:A:70:LYS:HZ3	1:A:70:LYS:CB	2.28	0.44
1:A:82:VAL:HG22	1:A:83:ASN:N	2.33	0.44
1:A:29:ASP:OD2	2:P:2:ALA:HB1	2.19	0.43
1:B:66:ILE:O	1:B:69:HIS:HB2	2.20	0.41
1:A:14:ARG:HG3	1:A:19:LEU:HD23	2.01	0.41
1:A:95:CYS:HA	1:B:99:PHE:HB3	2.02	0.41
1:B:14[1]:ARG:NH1	1:B:17:GLY:O	2.53	0.41
1:A:18:GLN:HE21	1:A:18:GLN:HB3	1.72	0.40

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	101/99 (102%)	101 (100%)	0	0	100	100
1	B	101/99 (102%)	101 (100%)	0	0	100	100
2	P	7/10 (70%)	7 (100%)	0	0	100	100
All	All	209/208 (100%)	209 (100%)	0	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	76/83 (92%)	74 (97%)	2 (3%)	40	44
1	B	77/83 (93%)	76 (99%)	1 (1%)	61	68
2	P	6/6 (100%)	6 (100%)	0	100	100
All	All	159/172 (92%)	156 (98%)	3 (2%)	50	56

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	67	CYS
1	A	70	LYS
1	B	24	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (9) such

sidechains are listed below:

Mol	Chain	Res	Type
1	A	2	GLN
1	A	18	GLN
1	A	25	ASN
1	A	61	GLN
1	A	69	HIS
1	A	88	ASN
1	A	98	ASN
1	B	18	GLN
1	B	98	ASN

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

2 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
3	ACT	A	501	-	3,3,3	2.61	1 (33%)	3,3,3	1.10	0
3	ACT	A	503	-	3,3,3	2.93	1 (33%)	3,3,3	1.11	0

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	503	ACT	O-C	4.97	1.43	1.22
3	A	501	ACT	O-C	4.42	1.41	1.22

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

EDS was not executed - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates

EDS was not executed - this section is therefore empty.

6.4 Ligands

EDS was not executed - this section is therefore empty.

6.5 Other polymers

EDS was not executed - this section is therefore empty.